



(19) **United States**

(12) **Patent Application Publication**

Jun et al.

(10) **Pub. No.: US 2001/0021267 A1**

(43) **Pub. Date: Sep. 13, 2001**

(54) **METHOD OF DETECTING DISSOLVE/FADE IN MPEG-COMPRESSED VIDEO ENVIRONMENT**

Publication Classification

(51) **Int. Cl.⁷** **G06K 9/00; G06K 9/36; G06K 9/46**

(52) **U.S. Cl.** **382/165; 382/232**

(57) **ABSTRACT**

There is provided a method of detecting dissolve/fade in an MPEG-compressed video environment, which includes the steps of: detecting a candidate sequence that is presumed to use a dissolve/fade editing effect according to shot transition detection in a video sequence; finding if spatio-temporal macro block type distribution that characteristically appears in a dissolve/fade sequence arises in the dissolve/fade candidate sequence, to judge if a scene transition by dissolve/fade was used in the detected dissolve/fade candidate sequence; and when the spatio-temporal macro block type distribution in the dissolve/fade sequence continuously appears in the dissolve/fade candidate sequence, comparing the length of the candidate sequence with a predetermined critical value and finally judging that the candidate sequence is a dissolve/fade sequence when its length is longer than the threshold.

(75) Inventors: **Sung Bae Jun**, Seoul (KR); **Kyoung Ro Yoon**, Seoul (KR)

Correspondence Address:
FLESHNER & KIM, LLP
P.O. Box 221200
Chantilly, VA 20153-1200 (US)

(73) Assignee: **LG Electronics Inc.**

(21) Appl. No.: **09/797,962**

(22) Filed: **Mar. 5, 2001**

(30) **Foreign Application Priority Data**

Mar. 7, 2000 (KR) 11334/2000

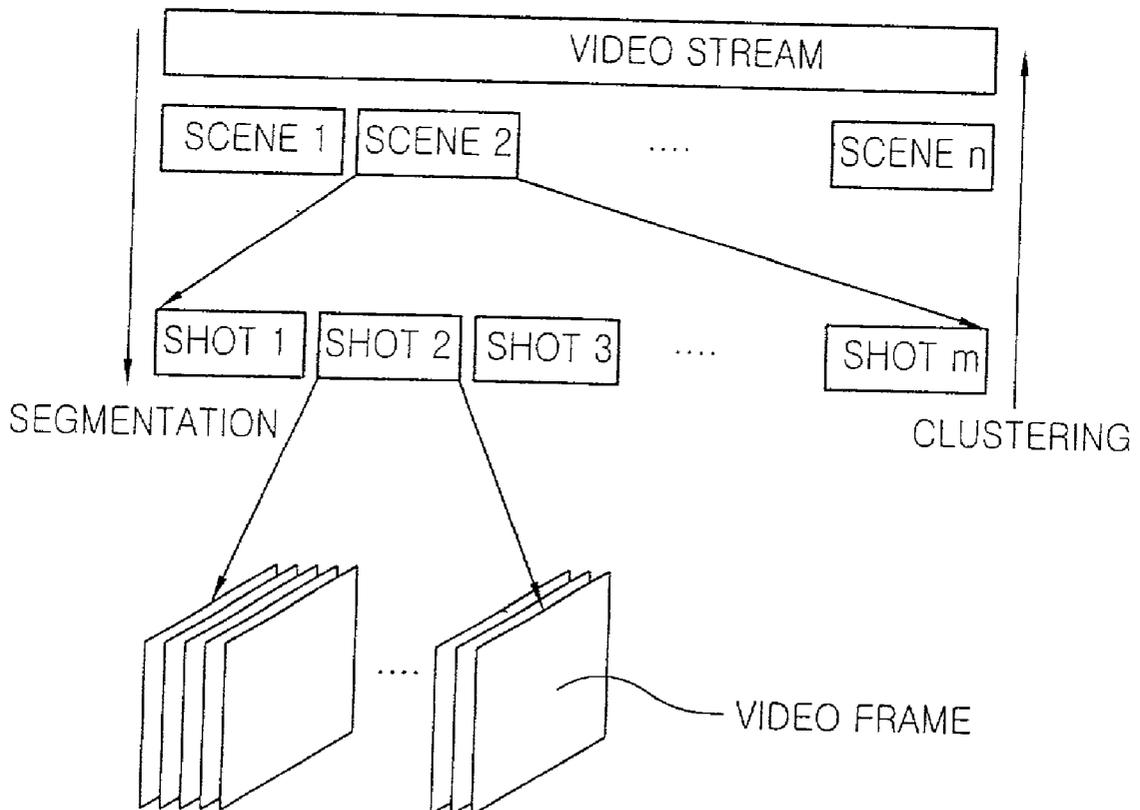


FIG. 1

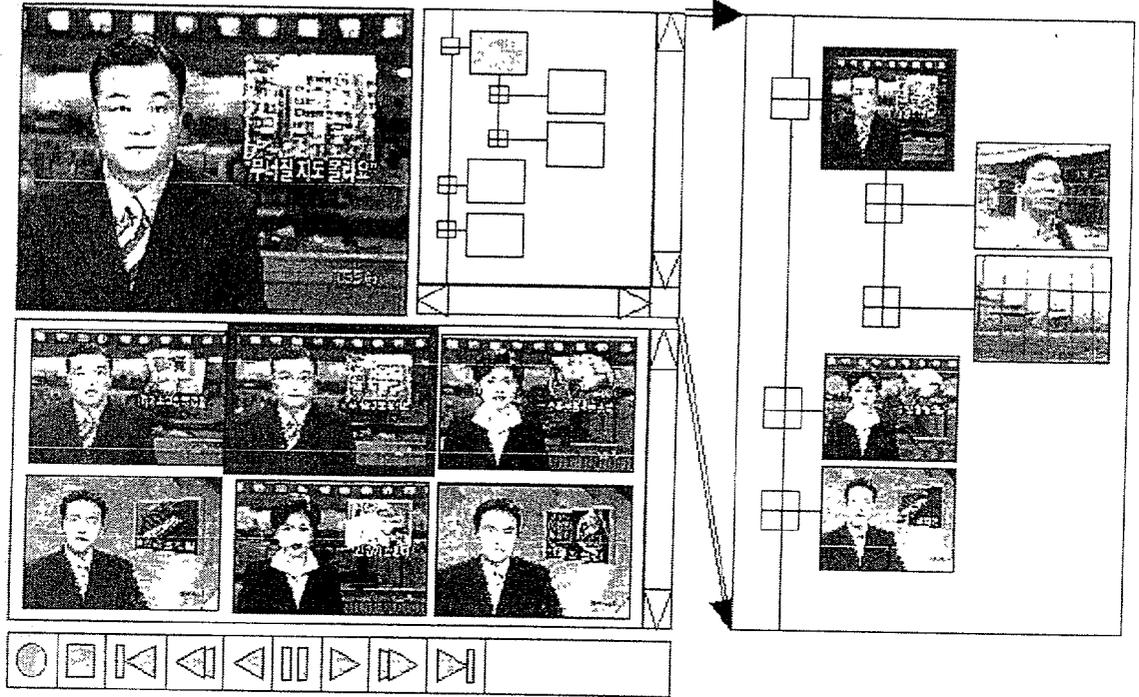


FIG. 2

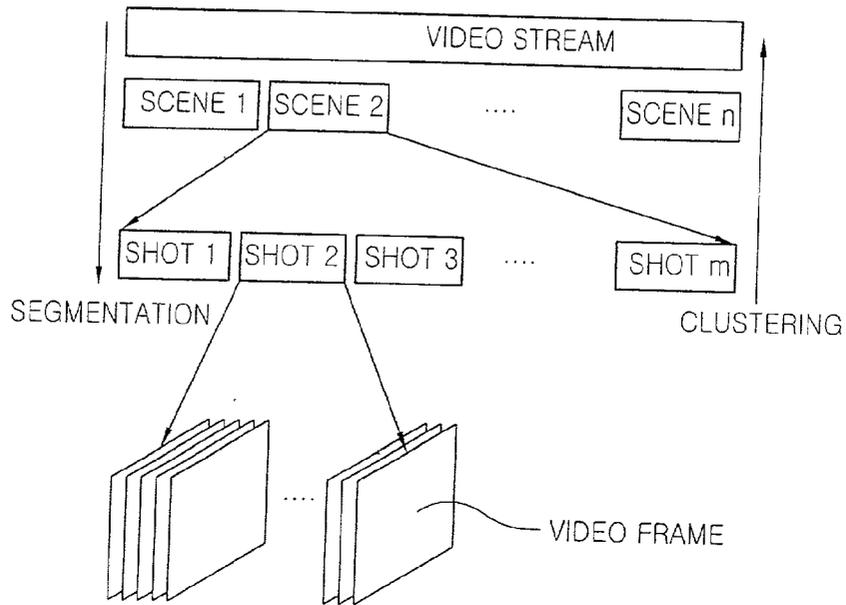


FIG. 3

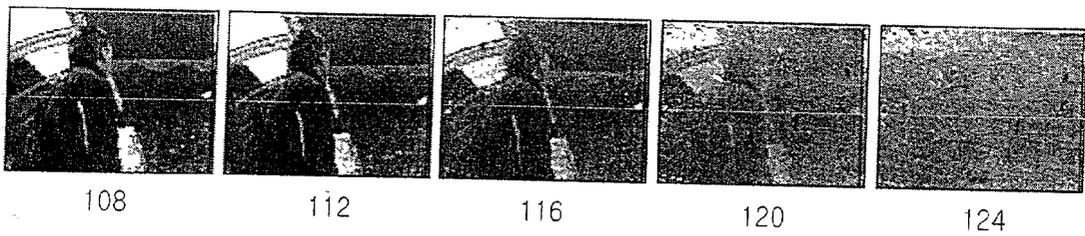


FIG. 4

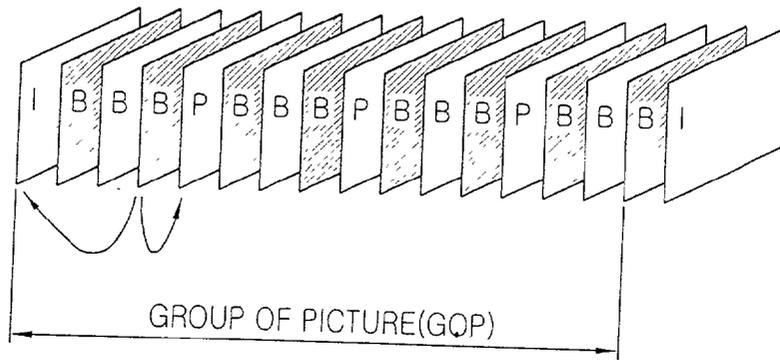


FIG. 5A

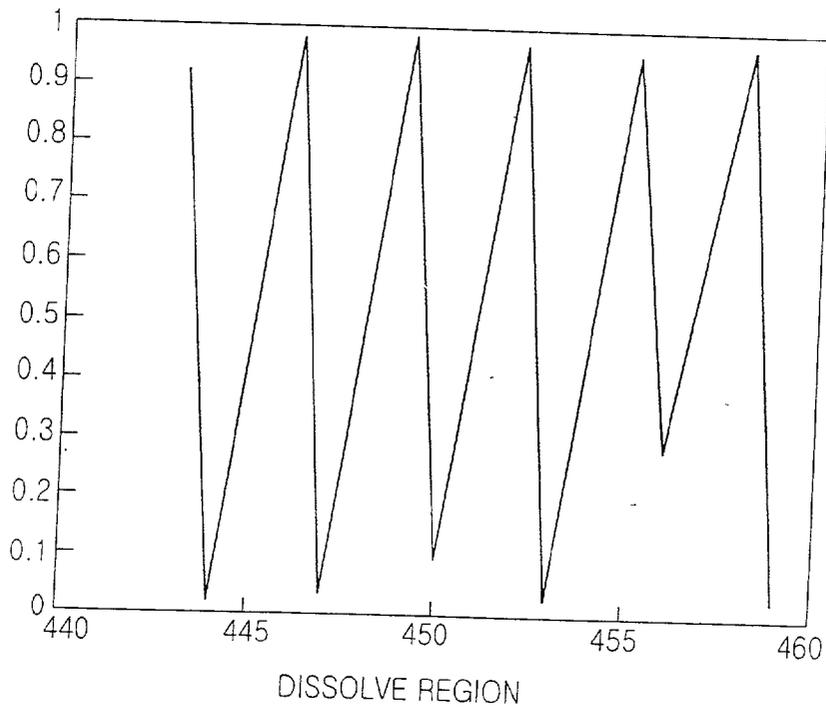


FIG. 5B

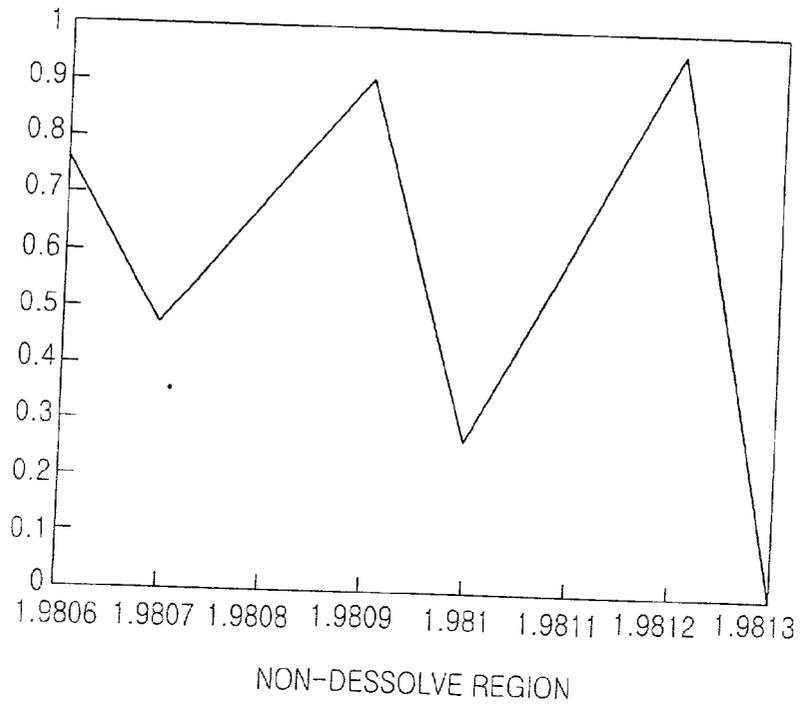
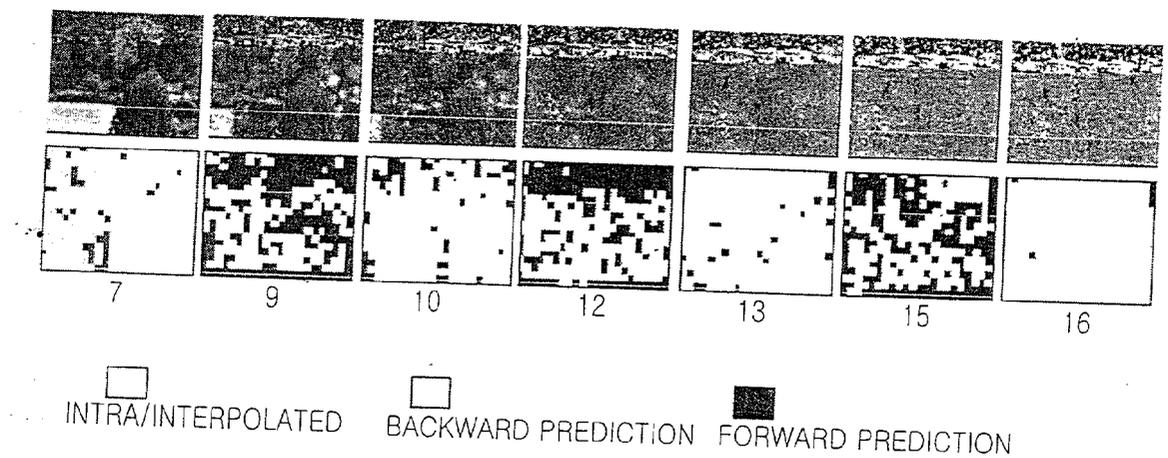


FIG. 6



METHOD OF DETECTING DISSOLVE/FADE IN MPEG-COMPRESSED VIDEO ENVIRONMENT

BACKGROUND OF THE INVENTION

[0001] 1. Field of the Invention

[0002] The present invention relates to a method of detecting dissolve/fade in an MPEG-compressed video environment, and more particularly, to a method of detecting a dissolve/fade sequence using spatio-temporal macro block type distribution in a compressed video environment, to effectively detect dissolve/fade in video streams.

[0003] 2. Description of the Related Art

[0004] To watch a desired video (moving picture such as movie, drama, news, documentary, etc.) through TV and video media, a user should watch the entire program at a fixed televising time. With the development in digital technology and image/video recognition techniques in recent years, however, users can search and browse a desired part of a desired video at a desired time. A basic technique for nonlinear browsing and searching includes a shot segmentation and a shot clustering. A variety of studies are being performed for the shot segmentation technique while researches with respect to the shot clustering technique are at the initial stage.

[0005] A shot is a sequence of video frames obtained by one camera without being interrupted/interruption. The shot is a basic unit for analyzing or constructing a video content. Video is generally configured of a connection of lots of shots and various video editing effects are used according to methods of connecting the shots. The video editing effects include an abrupt shot transition and a gradual shot transition. The abrupt shot transition is a technique that the current picture is abruptly changed into another picture. This abrupt shot transition is also called hard cut and being prevalently used. The gradual shot transition is a technique that a picture is gradually changed into another picture such as fade, dissolve, wipe and other special effects. Among these, the fade and dissolve are most frequently used.

[0006] Shot segmentation represents a process of extracting temporal information, such as frame numbers, of each shot of a video based on the transition detection.

[0007] There are many shot transition detection algorithms which of each can be categorized as three conventional methods for detecting the gradual shot transition more effectively. The first one is a twin comparison technique based on a color histogram difference between frames. This technique has erroneous detection and non-detection and slower performance speed because it is based on only the global color histogram difference between frames. The second method is a dissolve/fade detection technique based on the variance of global brightness distribution of frame. This technique uses brightness variation characteristic in I-frames and P-frames of a fade/dissolve sequence and also uses facts that a brightness variance graph has a parabolic form and very large difference between the maximum and minimum values and the editing effect of dissolve or fade lasts over several to tens frames. However, the brightness variance distribution used as a basis for detecting the dissolve/fade effect in this method frequently appears even in a sequence where dissolve/fade is not generated. Moreover, it is not arise in the sequence where the dissolve/fade is generated in many cases.

[0008] The third method is a dissolve/fade detecting technique based on edge distribution in an image according to an edge detection algorithm and analysis of moving picture characteristic of the detected edge. This method passes through a preprocessing step of detecting edges from image data, a step of dividing the detected edges into entering edges and exiting edges using the moving picture characteristic and calculating an edge variation rate on the basis of the divided edges, and a post-processing step of classifying editing effects using spatio-temporal distribution of the entering edges and exiting edges, to detect the editing effects of hard cut, dissolve, fade and wipe. However, this method has very a slow performance speed because most images must be actually decoded basically and the edge detection operation requires relatively long period of time.

SUMMARY OF THE INVENTION

[0009] It is, therefore, an object of the present invention to provide a method of detecting dissolve/fade in an MPEG-compressed video environment, which rapidly and accurately detects a sequence where dissolve/fade is generated based on spatio-temporal macro block type distribution in a video compression domain using bi-directional prediction between frames.

[0010] To accomplish the object of the present invention, there is provided a method of detecting dissolve/fade in an MPEG-compressed video environment, comprising the steps of: detecting a candidate sequence that is presumed to use a dissolve/fade editing effect according to shot transition detection in a video sequence; finding if spatio-temporal macro block type distribution that characteristically appears in a dissolve/fade sequence arises in the dissolve/fade candidate sequence, to judge if a scene transition by dissolve/fade was used in the detected dissolve/fade candidate sequence, and when the spatio-temporal macro block type distribution in the dissolve/fade sequence continuously appears in the dissolve/fade candidate sequence, comparing the length of the candidate sequence with a predetermined critical value and finally judging that the candidate sequence is a dissolve/fade sequence when its length is longer than the critical value.

BRIEF DESCRIPTION OF THE DRAWINGS

[0011] FIG. 1 shows a non-linear video browsing interface according to an embodiment of the present invention;

[0012] FIG. 2 shows a relationship between the shot segmentation and shot clustering according to the present invention;

[0013] FIG. 3 shows an example of shot transition by dissolve in a video sequence in accordance with the present invention;

[0014] FIG. 4 shows the structure of GOP in an MPEG video sequence according to the present invention;

[0015] FIGS. 5A and 5B are graphs showing forward prediction rates in a dissolve/fade sequence and a non-dissolve/fade sequence in an MPEG video sequence according to the present invention, respectively; and

[0016] FIG. 6 shows distributions by macro block types in B-frames adjacent to anchor frames in a dissolve sequence according to the present invention.

DETAILED DESCRIPTION OF THE PREFERRED EMBODIMENT

[0017] Reference will now be made in detail to the preferred embodiments of the present invention, examples of which are illustrated in the accompanying drawings.

[0018] FIG. 1 shows a non-linear video browsing interface according to an embodiment of the present invention. This interface is very useful in digital video browsing because a user can easily access only a desired part of video by searching main parts thereof using the interface without watching the whole part of the video. The most essential technique for video browsing includes the shot segmentation and shot clustering. The relationship between the shot segmentation and shot clustering is explained below with reference to FIG. 2.

[0019] FIG. 2 shows the relationship between the shot segmentation and shot clustering according to the present invention. Referring to FIG. 2, a video stream is configured of a connection of scenes logically constructed, each scene being configured of a connection of a lot of shots. The shot segmentation is a technique for dividing the video stream into individual shots. The shot clustering is a technique of grouping similar shots on the basis of similarity of time/image/motion/audio to construct a video structure in unit of logical scene.

[0020] A video editing effect is classified based on methods of connecting the shots. The editing effect includes the abrupt transition, that is, hard cut, and gradual transition such as dissolve, fade, wipe and other special effects. The dissolve and fade are most frequency used for gradual connection of two shots or scenes in video edition. The dissolve is a technique that two scenes are overlapped each other to be gradually changed from one scene to the other scene. The fade is a technique that a scene fades out or fades in, being gradually changed into another scene.

[0021] There is described below a method of detecting fade and dissolve using spatio-temporal macro block type distribution of a video MPEG-compressed according to bi-directional prediction between frames with reference to FIGS. 3 to 6. FIG. 3 shows an example of shot transition by dissolve in a video sequence in accordance with the present invention. Referring to FIG. 3, as the video sequence proceeds, transition from one scene 108 into another scene 124 occurs, the two scenes being overlapped each other.

[0022] When the shot transition sequence detected using dissolve/fade is analyzed in the video, it has the following characteristics. Firstly, there is a considerable difference between color distributions of the starting scene and ending scene of the dissolve/fade. Secondly, the dissolve/fade generally lasts for more than several frames. Thirdly, the first scene gradually becomes dim and the second scene gradually becomes bright in the dissolve/fade. Finally, pixels that become dim and pixels that become bright spatially widely distribute. On the basis of these characteristics, the present invention realizes an algorithm for effectively detecting the dissolve/fade using spatio-temporal macro block type distribution characteristic in B-frames that simultaneously use bi-directional prediction in the compression domain.

[0023] A procedure for realizing the algorithm is as follows.

[0024] First of all, a candidate sequence that is presumed to use the dissolve/fade technique is detected from a video sequence through shot transition detection. This candidate sequence is judged to be a sequence where the dissolve/fade was generated when a color histogram difference between the first frame and the last frame of a scene where the dissolve/fade is detected is larger than a predetermined threshold. This can be represented by the following expression.

$$\text{HistDiff}(f_b, f_e) > \tau_{\text{color}} \quad (1)$$

[0025] Where f_b is the starting frame of dissolve/fade scene, f_e is the ending frame of the dissolve/fade scene, $\text{HistDiff}(f_b, f_e)$ is the color histogram difference between f_b and f_e , and τ_{color} is the predetermined threshold for judgement of the generation of shot transition based on the color histogram difference.

[0026] The candidate sequence can be detected using not only a method of detecting the shot transition based on global color distribution difference between frames but also a method based on the spatio-temporal macro block distribution and a method based on spatio-temporal edge distribution and variation form characteristic. There are explained methods of detecting the frames f_b, f_e serving as a base of color distribution comparison in the method using the color distribution difference. First of all, there is a method of selecting a frame of one-step interval from a reference frame. Another method is to detect I-frames as the candidate sequence ($[f_b, f_e]$) which uses only intra coded blocks in video CODEC such as H.xxx or MPEG.

[0027] It is judged if there is a hard cut in the dissolve/fade candidate sequence ($[f_b, f_e]$) detected as above. This can improve accuracy in the dissolve/fade detection algorithm. The hard cut is detected through a variety of methods including a technique using an image difference between two frames according to global color distribution difference based on color histogram, a technique using spatio-temporal macro block distribution and its variation characteristic and a technique using spatio-temporal motion vector characteristic, spatio-temporal edge distribution through edge detection and its variation characteristic.

[0028] In case where it is judged that there is no hard cut, it is found if the dissolve/fade editing effect was used in the detected dissolve/fade candidate sequence ($[f_b, f_e]$) based on existence of spatio-temporal macro block type distribution that characteristically appears in dissolve/fade sequence. Checking of the spatio-temporal macro block type distribution is performed on B-frames that are coded using bi-directional prediction between frames. The selected B-frames are adjacent to anchor frames in the candidate sequence ($[f_b, f_e]$). The anchor frames are I-frames or P-frames serving as a base of motion prediction/compensation between frames. The above-described B-frames, I-frames and P-frames are explained below in detail with reference to FIG. 4.

[0029] FIG. 4 shows a structure of GOP (Group of Picture) in an MPEG video sequence according to the present invention. The GOP is one of specific MPEG video sequences. In FIG. 4, black-colored frames represent B-frames adjacent to anchor frames and these frames are accessed to detect the dissolve/fade in minimal decoding domain. The anchor frames serve as a base frame for

prediction/motion compensation between frames, and the B-frame has two anchor frames all the time. In the present invention, only the B-frames adjacent to anchor frames are accessed without accessing all of B-frames in order to minimize decoding, the dissolve/fade can be accurately detected.

[0030] To obtain a varied dissolve/fade candidate sequence $([f_b, f_e])$ in the candidate sequence $([f_b, f_e])$ that satisfies the spatio-temporal macro block distribution characteristic of the dissolve/fade sequence.

[0031] The larger value between forward prediction rate and backward prediction rate is larger than a predetermined critical value. This is represented by the following expressions.

$$\text{Max}(M_{\text{fwd}}/(M_{\text{fwd}}+M_{\text{bwd}}), M_{\text{bwd}}/(M_{\text{fwd}}+M_{\text{bwd}})) > \tau_t \quad (2)$$

$$(\text{if } M_{\text{fwd}}+M_{\text{bwd}} \neq 0) \text{ SpatDist}(\text{MinType}(M_{\text{fwd}}, M_{\text{bwd}})) > \tau_s \quad (3)$$

$$(\text{if } M_{\text{fwd}}, M_{\text{bwd}} = 0) \text{ MinType}(M_x, M_y) = X(\text{if } M_x < M_y) \quad (4)$$

$$\text{MinType}(M_x, M_y) = Y(\text{if } M_x > M_y) \quad (5)$$

[0032] Where M_{fwd} is the number of forward prediction macro blocks of frame, M_{bwd} is the number of backward prediction macro blocks of frame, τ_t is the critical value for the ratio of forward prediction and backward prediction, $M_{\text{fwd}}/(M_{\text{fwd}}+M_{\text{bwd}})$ is the forward prediction rate, $M_{\text{bwd}}/(M_{\text{fwd}}+M_{\text{bwd}})$ is the backward prediction rate, $\text{SpatDist}(A)$ is spatial distribution measurement function of macro blocks whose type is A in an image, and τ_s is a critical value for the spatial distribution measurement of macro blocks. If a B-frame in the candidate sequence satisfies (2) and (3), the B-frame will be set to 1.

[0033] After the aforementioned procedure, there is detected the candidate sequence $([f_b, f_e])$ having the maximum length among runs set to 1 among the B-frames adjacent to the anchor frames within the obtained sequence $([f_b, f_e])$.

[0034] When the larger value between the forward prediction rate and backward prediction rate is larger than the specific threshold (expression (2)), the forward or backward prediction rate is considerably high in the B-frames adjacent to the anchor frames in the dissolve/fade sequence. The expressions model that this phenomenon continuously appears in the dissolve sequence. Moreover, the above expressions use characteristics that macro block prediction rate is much higher and appears continuously in the dissolve/fade sequence although it is general that more macro blocks are predicted from closer anchor frames in the B-frames. These characteristics are represented by graphs of **FIGS. 5A and 5B**. **FIGS. 5A and 5B** are graphs showing forward prediction rates in a dissolve/fade sequence and a non-dissolve/fade sequence in an MPEG video sequence according to the present invention, respectively.

[0035] The expression (3) represents the forward prediction macro blocks and backward prediction macro blocks are globally scattered in the spatial domain. The expression is for reducing erroneous detection rate in the entire algorithm.

[0036] The spatial distribution measurement function is a method of judging how much a specific type macro block is spatially globally distributed in an image. As an example,

$\text{SpatDist}(A)$ for measuring spatial distribution of A-type macro block can be represented by the following expression.

$$\text{SpatDist}(A) = C_A / T_A \quad (6)$$

[0037] Where C_A is the total number of connected components on the basis of type A, and T_A is the total number of A-type macro blocks in an image.

[0038] **FIG. 6** shows distributions of macro block types in B-frames adjacent to the anchor frame in a dissolve sequence according to the present invention. The function that induces the forward prediction macro block and backward prediction macro block to globally scattered in the spatial domain is the spatial distribution measurement function that judges that macro blocks of two types globally scattered in the image. This function has higher value as the macro blocks of two types are more globally distributed. The spatial distribution measurement function is decided by a value obtained by dividing the number of connected components of a specific type macro block by the number of a specific type macro block in an image.

[0039] In the analysis of the spatial distribution measurement, a macro block type in smaller numbers is selected but, if required, a macro block type in larger numbers can be selected for checking the spatial distribution.

[0040] After passing through the above procedures, the dissolve/fade detecting algorithm using the spatio-temporal macro block type distribution applies time constraint in order to judge if corresponding candidate sequence is an actual scene transition sequence according to dissolve/fade. That is, the corresponding sequence is judged to be the scene transition sequence by dissolve/fade when the spatio-temporal characteristic of the macro block type distribution in B-frames continuously appears for a predetermined period of time in the dissolve/fade sequence continuously appears for a predetermined period of time. On the other hand, it is judged that the corresponding sequence is not the scene transition sequence by dissolve/fade when it is not. The length of the dissolve/fade candidate sequence $([f_b, f_e])$ or $([f_e, f_b])$ having the maximum length, which was detected through the above procedure, is compared with a specific threshold (τ_t). When the length is larger than the threshold value, this sequence $([f_b, f_e])$ or $([f_e, f_b])$ is decided as the dissolve/fade sequence, thereby detecting dissolve/fade. This is represented by the following expression.

$$[e'-b] = \tau_t \quad (7)$$

[0041] where τ_t is a modeled duration. Furthermore, when variance of colors of the first scene of the dissolve/fade candidate sequence obtained through the above procedures is lower than a predetermined critical value, the sequence is judged to be fade-in. When variance of colors of its last scene is lower than the critical value, the sequence is judged to be fade-out. The sequence is judged to be dissolve in other cases. Accordingly, the dissolve and fade can be discriminated from each other by the following expressions.

$$\text{[0042]} \quad \text{if ColorDist}(f_{\text{start}}) < \tau_d \text{ then Fade-In}$$

$$\text{[0043]} \quad \text{else if ColorDist}(f_{\text{end}}) < \tau_d \text{ then Fade-Out}$$

$$\text{[0044]} \quad \text{else dissolve}$$

[0045] $\text{ColorDist}(f_i)$ is a measure for indicating how various colors compose the image of frame f_i and it can be applied to only pixels that are sampled on the specific basis.

In the above expressions, τ_d is a threshold for deciding fade-in and fade-out, f_{start} is the starting point of time of dissolve/fade, and f_{end} is the ending point of time of dissolve/fade. f_{start} can use f_b or f_e , and f_{end} can use f_c or f_e . The above expressions use a characteristic that a picture starts from a simple scene in fade-in and the picture becomes simple in fade-out.

[0046] As distinguished from the conventional algorithm of detecting dissolve/fade, the present invention detects the dissolve/fade using the spatio-temporal macro block type distribution and its variation form in B-frames that compensate motions and perform bi-directional prediction in minimal decoding domain.

[0047] The dissolve/fade detecting method of the invention has a performance speed higher than the conventional algorithm because its processing is carried out in the minimal decoding domain. Furthermore, it is robust against fast camera motions or large motion information of a large object. Moreover, the present invention provides an algorithm capable of rapidly and accurately detecting fade/dissolve effects widely used among the gradual shot transition in the shot segmentation field. This algorithm uses basic features used in the shot segmentation algorithm so that it can be easily combined with the conventional shot segmentation algorithm. Also, it can be used as a basic input for shot clustering.

[0048] Although specific embodiments including the preferred embodiment have been illustrated and described, it will be obvious to those skilled in the art that various modifications may be made without departing from the spirit and scope of the present invention, which is intended to be limited solely by the appended claims.

What is claimed is:

1. A method of detecting dissolve/fade in an MPEG-compressed video environment, comprising the steps of:

detecting a candidate sequence that is presumed to use a dissolve/fade editing effect according to shot transition detection in a video sequence;

finding if spatio-temporal macro block type distribution that characteristically appears in a dissolve/fade sequence arises in the dissolve/fade candidate sequence, to judge if a scene transition by dissolve/fade was used in the detected dissolve/fade candidate sequence; and

when the spatio-temporal macro block type distribution in the dissolve/fade sequence continuously appears in the dissolve/fade candidate sequence, comparing the length of the candidate sequence with a predetermined critical value and finally judging that the candidate sequence is a dissolve/fade sequence when its length is longer than the critical value.

2. The method as claimed in claim 1, wherein it is judged that there is a hard cut in the detected candidate sequence and, only when there is no hard cut, the process goes to the next step.

3. The method as claimed in claim 1, wherein the candidate sequence is judged to be the dissolve/fade sequence when a color histogram difference between the first frame and the last frame of a scene from which dissolve/fade is detected is larger than a predetermined threshold.

4. The method as claimed in claim 3, wherein frames serving as a base for comparison of global color distributions are selected by a method of selecting a frame of one-step interval from a reference frame, or selecting I-frames coded using only intra-coded blocks as candidate frames.

5. The method as claimed in claim 1, wherein the step of detecting the dissolve/fade candidate sequence is performed by a technique of using an image difference between two frames by using a difference in color histogram, a technique of using spatio-temporal macro block distribution, a technique of using spatio-temporal motion vector distribution, or a technique of using spatio-temporal edge distribution and its variation characteristic.

6. The method as claimed in claim 2, wherein the hard cut is detected by a method of using an image difference between two frames by using a difference in color histogram based global color distribution, a method of using spatio-temporal macro block distribution, a method of using spatio-temporal motion vector distribution, or a method of using spatio-temporal edge distribution and its variation form characteristics.

7. The method as claimed in claim 1, wherein the step of judging if the dissolve/fade editing effect was used in the candidate sequence using the spatio-temporal macro block type distribution uses spatio-temporal macro block type distribution and its variation characteristics in B-frames that simultaneously use bi-directional prediction in compression domain.

8. The method as claimed in claim 7, wherein the selected B-frames are adjacent to anchor frames on the basis of the anchor frames in the candidate sequence.

9. The method as claimed in claim 8, wherein the anchor frames are I-frames or P-frames serving as a base for motion prediction/compensation between frames.

10. The method as claimed in claim 7, wherein the step of judging if the dissolve/fade editing effect was used in the candidate sequence using the spatio-temporal macro block type distribution comprises the steps of:

setting B-frames whose macro block type distribution satisfies "B-frame macro block type characteristic in a dissolve/fade sequence" among the B-frames adjacent to the anchor frames in the dissolve/fade candidate sequence to 1 and setting other B-frames to 0; and

obtaining a run having a maximum length among the runs set to 1.

11. The method as claimed in claim 10, wherein "B-frame macro block type characteristic in a dissolve/fade sequence" is that the sum of the number of forward prediction macro blocks and the number of backward prediction macro blocks in corresponding B-frame is not equal to 0 and larger value between forward prediction rate and backward prediction rate is larger than a threshold.

12. The method as claimed in claim 10, wherein "B-frame macro block type characteristic in a dissolve/fade sequence" is that one of the number of forward prediction macro blocks and the number of backward prediction macro blocks in corresponding B-frame is 0 or both are not equal to 0, and the forward prediction macro blocks and backward prediction macro blocks are globally scattered in the spatial domain.

13. The method as claimed in claim 10, wherein a function representing the spatial distribution inputs the num-

ber of connected components of a specific type macro block and the number of specific type macro blocks in an image, and it is decided by a value obtained by dividing the inputted number of connected components by the inputted number of the specific type macro blocks in the image.

14. The method as claimed in claim 12, wherein a function that induces the forward prediction macro blocks and backward prediction macro blocks to be globally scattered in the spatial domain is a function (spatial distribution measurement function) of judging that macro blocks of two types are globally scattered in the image of an image type macro block, the function having a higher value as the macro blocks of two types are more globally scattered, and it is judged that corresponding B-frame satisfies "B-frame macro block type characteristic in a dissolve/fade sequence" when the result of the function exceeds a threshold for the spatial distribution of the macro blocks.

15. The method as claimed in claim 12, wherein the spatial distribution measurement function selects a type in

smaller numbers among the forward macro blocks and backward macro blocks to use it as an input, or selects a type in larger numbers among them to use it as an input.

16. The method as claimed in claim 1, wherein the dissolve/fade candidate sequence is judged to be fade-in when variance of colors for the first scene in the candidate sequence is lower than a predetermined threshold, it is judged to be fadeout when variance of colors for the last scene is lower than a threshold for discriminating fade-in and fade-out from each other, and it is judged to be dissolve in other cases.

17. The method as claimed in claim 16, wherein the variance of colors is based on diversity of colors constructing pixels in an image while brightness is based on diversity of colors constructing sampled pixels among pixels in an image.

* * * * *